



Best Available Copy

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/785,560	02/24/2004	Christopher J. C. Burges	MS305553.1/MSFTP561US	8133
27195 7590 01/24/2008 AMIN. TUROCY & CALVIN, LLP 24TH FLOOR, NATIONAL CITY CENTER 1900 EAST NINTH STREET CLEVELAND, OH 44114			EXAMINER COLUCCI, MICHAEL C	
			ART UNIT 2626	PAPER NUMBER
			NOTIFICATION DATE 01/24/2008	DELIVERY MODE ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket1@thepatentattorneys.com  
hholmes@thepatentattorneys.com  
osteuball@thepatentattorneys.com

## Office Action Summary

Application No.

10/785,560

Applicant(s)

BURGES ET AL.

Examiner

Michael C. Colucci

Art Unit

2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☒ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-37 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-9, 13, 15-24 and 31-37 is/are rejected.
- 7) ☒ Claim(s) 10-12, 14 and 25-30 is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 24 February 2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
  - ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
  - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO/SB/08)  
Paper No(s)/Mail Date See Continuation Sheet.
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date. \_\_\_\_.
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: \_\_\_\_.

Continuation of Attachment(s) 3). Information Disclosure Statement(s) (PTO/SB/08), Paper No(s)/Mail Date :7/12/2007, 5/26/2005, 6/14/2004.

## **DETAILED ACTION**

### ***Response to Arguments***

1. Applicant's arguments with respect to claims 1-37 have been considered but are moot in view of the new ground(s) of rejection.

### ***Allowable Subject Matter***

2. Claims 10-12, 14, and 25-30 objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

### ***Claim Rejections - 35 USC § 112***

3. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

4. Claims 1, 17, and 18 rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. Within the 4<sup>th</sup> limitation of claims 1, 17, and 18: "wherein a gap is a temporal space between two adjacent fingerprints that exceeds a predetermined threshold when fingerprints within a set of clusters are placed in

Art Unit: 2626

sequential temporal order". The specific definition of a "gap" is not disclosed within the specification. Gap as disclosed in the specification is relevant to a gap in an ordering of fingerprints with no relevance to a threshold.

5. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

6. Claim 25 rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Applicant fails to define " $t_0 \square 0$ " and " $t_{n+1} \square 1$ ".

### ***Claim Rejections - 35 USC § 103***

7. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

8. Claims 1-5, 7, 13, 17-19, and 31-33 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang).

Re claims 1, 17, and 18 Cooper teaches a system for summarizing audio information, comprising:

an analyzer to convert audio into frames ([0052] & fig. 7);

a fingerprinting component to convert the frames into fingerprints, each fingerprint based in part on a plurality of frames ([0052] & fig. 7);

a similarity detector to compute similarities between fingerprints, the similarity detector comprising a clustering function ([0021]), the clustering function producing one or more sets of clusters of fingerprints based upon all fingerprints within a set of clusters meeting an initial threshold indicative of similarity ([0045] & fig. 7);

However, Cooper fails to teach a heuristic module to generate a thumbnail of the audio file (Zhang [0011]) from a set of clusters that has at least two gaps between fingerprints, wherein a gap is a temporal space between two adjacent fingerprints that exceeds a predetermined threshold when fingerprints within a set of clusters are placed in sequential temporal order (Zhang [0034] & Fig. 4).

NOTE: A thumbnail is construed as a summarization of an audio file. Also, Zhang teaches periods between vocal portions of an audio track.

Zhang teaches the detection of an occurrence of an increase in energy within the audio track, where there is an extraction of a second portion of the audio track corresponding to the detected increase in energy. Zhang also teaches within figure 4, the box 400 indicating an interlude period of the audio track, while the line 402 denotes the start of singing voice following the interlude.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention generating a thumbnail of an audio file with several gaps between fingerprints. Having several gaps/periods between fingerprints would allow for a more

Art Unit: 2626

robust signal, where no overlapping of data will occur during the analysis of segments/frames.

Re claim 2, Cooper particular fails to teach the system of claim 1, the heuristic module comprising at least one of an energy component and a flatness component (Zhang [0045]) in order to help determine a suitable segment of audio for the thumbnail (Zhang [0034] & Fig. 4;).

Zhang teaches the appearance of low level energy minimums after a relatively long period of continuous high energy values which may also indicate the start of a singing voice. Zhang also teaches within figure 4, the box 400 indicating an interlude period of the audio track, while the line 402 denotes the start of singing voice following the interlude.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention recognizing energy components would allow for the determination of a chorus through analysis of the frequency domain of the audio signal, which would allow for verifying that a chorus exists where the fingerprints fail to find a chorus. By recognizing flat energy components, audio portions such as chorus can be recognized and compared for repetition to determine a true chorus, where a low energy area is typically indicative of an instrumental segment of a song without voice.

Re claim 3, Cooper teaches the system of claim 2, the heuristic module is employed to automatically select voiced choruses over instrumental portions ([0048]).

Re claim 4, Cooper fails to particularly teach the system of claim 2, the energy component and the flatness component are employed when the fingerprints do not result in finding a suitable chorus (Zhang [0045] & Fig. 4).

Zhang teaches the appearance of low level energy minimums after a relatively long period of continuous high energy values may also indicate the start of a singing voice. Zhang also teaches a zero crossing rate in addition to the length of a segment of audio that can be used to distinguish between chorus versus verse or interlude.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention energy and flatness components used when fingerprints do not result in finding a suitable chorus. Recognizing energy components in addition to flatness components would allow for the determination of a chorus through analysis of the frequency domain of the audio signal, which would allow for verifying that a chorus exists where the fingerprints fail to find a chorus. By eliminating all zero (flat) energy components, audio portions such as chorus can be recognized and compared for repetition to determine a true chorus.

Re claim 5, Cooper fails to particularly teach the system of claim 1, further comprising a component to remove silence at the beginning and end of an audio clip via an energy-based threshold (Zhang [0053]).

Zhang teaches an audio thumbnail for each track stored, where the musical piece can be viewed by application of the respective first and second pointers and



Art Unit: 2626

durations of time to respectively play the musical piece starting where the first pointer directs, for a first duration, and then starting where the second pointer directs, for a second duration. In such a manner, an excerpt of a song or other audio composition can be previewed for an ultimate listening or purchasing decision (Zhang [0053]).

Zhang also teaches detecting the occurrence of a first content feature or characteristic within the audio track and a pointer that is mapped to the location on the audio track where the occurrence of a highlight, such as the largest increase in temporal energy within the audio track, has been detected, where durations of time are set.

NOTE: If singing is only preserved, silence at the beginning, end, or middle will not be stored will thus be removed from the audio thumbnail.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a component to remove silence at the beginning and end of an audio-clip via an energy based threshold. Removing silent periods prior to generate a thumbnail would aid in a faster summary of the recognizable part of an audio clip or song that a user is trying to identify.

Re claim 7, Cooper teaches the system of claim 1, the analyzer computes a set of spectral magnitudes for an audio frame ([0020]).

Re claim 13, Cooper teaches a mean spectral quality ([0031])

However, Cooper fails to teach the system of claim 1, the heuristic component selects the set of clusters from which to generate the audio thumbnail based upon at

Art Unit: 2626

least one of a mean spectral quality value determined for the set of clusters or a cluster spread quality value determined for the set of clusters (Zhang [0011]).

Zhang teaches the detection of an occurrence of an increase in energy within the audio track and the extraction of a second portion of the audio track corresponding to the detected increase in energy. Zhang also teaches combining the extracted first and second portions of the audio track into an audio thumbnail of the audio track.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention selecting clusters to generate an audio thumbnail based on a mean spectral quality. Using a mean spectral quality would allow for an accurate audio thumbnail being that a chorus in a song repeats nearly identically each time, where the average/mean value of cluster set would be acceptable.

Re claim 19, Cooper teaches the method of claim 18, clustering fingerprints within a set of clusters into fingerprint clusters based upon the gaps (Cooper teaches portions of a song segmented into chorus, verse, intro, and bridge. Clusters, which contain similar fingerprints are construed to be separated by gaps or portions of a song that are not similar to the fingerprints within the group itself; fig. 7).

Re claim 31, Cooper fails to teach the method of claim 18, the creating further comprising determining a cluster by determining a longest section of audio within an audio file that repeats in the audio file (Zhang [0049] & Fig. 4 and 8).

Art Unit: 2626

Zhang teaches singing of the main melody represented in part B at 802 of figure 8. Zhang teaches that part C at 804 consists of another interlude and another paragraph of singing and part D at 806 shows the main melody being repeated twice. Zhang illustrates that the singing of the main melody at 802 and 806 can be seen as having a higher energy.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention determining a longest section of audio that repeats. Determining the longest repeating section in an audio file would aid in determining the chorus of a song, where the chorus would typically be the only repeating portion that would have higher energy than any instrumental portion present in a song.

Re claim 32, Cooper fails to teach the method of claim 18, the creating further comprising at least one of:

rejecting clusters that are close to a beginning or end of a song (Zhang [0049] & Fig. 4 and 8);

rejecting clusters for which energy falls below a threshold for any fingerprint in a predetermined window (Zhang [0050] & Fig. 7 items 706, 708);

selecting a fingerprint having a highest average spectral flatness measure in the predetermined window (Zhang 0034] & Fig. 4).

NOTE: For prior art purposes, a spectrally flat region is construed to both functionally equivalent and effective to the region 400 demonstrated by Zhang, where

Art Unit: 2626

the system can detect the specific frame with either high spectral zero crossing or low/flat zero crossing.

Zhang teaches a zero crossing rate when no speech is present in a signal prior to the start of sing/speaking, Where the zero crossing rate indicates that the signal is barely changing. Zhang also teaches singing of the main melody represented in part B at 802 of figure 8. Zhang teaches that part C at 804 consists of another interlude and another paragraph of singing and part D at 806 shows the main melody being repeated twice. Zhang illustrates that the singing of the main melody at 802 and 806 can be seen as having a higher energy. Zhang also teaches a magnitude of the increase compared to a predetermined threshold at step 708 as selected from control parameter storage 112 or as a default. If the compared increase in temporal energy exceeds the threshold, the compared increase and the location on the audio track corresponding to the beginning of the right window 902 are retained as indicative of a highlight on the audio track. In the figure, the box 400 indicates an interlude period of the audio track, while the line 402 denotes the start of singing voice following the interlude, at which point the relative increase in zero crossing rate value variances can be seen.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention rejecting clusters close to the beginning or end of a son, with energy below a threshold, having a highest average spectral flatness. Having energy below a threshold with spectral flatness will aid in indicating where there is no singing/speaking in a song typically at the beginning or end where an instrumental portion is present with lower energy levels and low zero crossing rate.

Re claim 33, Cooper teaches the method of claim 18, the creating further comprising generating a thumbnail by specifying time offsets in an audio file (Fig. 7).

**9. Claim 6 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Sweet US 6763136 B1 (hereinafter Sweet).**

Re claim 6, Cooper in view of Zhang fail to teach the system of claim 1, the fingerprint component further comprising a normalization component, such that an average Euclidean distance from the each fingerprint to other fingerprints for an audio clip is one. (Sweet col 5 line 25-35).

NOTE: For prior art purposes, a vector is construed to be a representation of a data in a frame in vector form that is functionally equivalent to a fingerprint.

Sweet teaches a normalized Euclidean Distance representative of the normalized average distance between a pair of spectral vectors that ranges in value between zero and one.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a value of one corresponding to an average Euclidean distance from each fingerprint to other fingerprints. Using an average Euclidean distance that is normalized to one would aid in determining allowing for another method to determine

Art Unit: 2626

similarity relative to distance between segments if frequency (energy) and time (fingerprint) domain selections were to fail.

**10. Claim 8 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Petkovic et al, US 6185527 (hereinafter Petkovic).**

Re claim 8, Cooper in view of Zhang fail to teach the system of claim 7, for each frame, a mean, normalized energy  $E$  is computed by dividing a mean energy per frequency component within the frame by the average of that quantity over frames in an audio file. (Petkovic col 10 line 37-45).

Petkovic teaches that calculated audio features are statistically normalized and the normalized version of a measured audio feature is the quotient of the difference between the measured audio feature and the mean value of that feature over all segments, and the standard deviation of that feature for all segments.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention normalization of energy through the division of an average energy component by the average of that quantity over frames in an audio file. Normalization through would aid in scaling data to a proper form to ensure that the current data in a frame is relevantly scaled against the average of the previous data.

Art Unit: 2626

**11. Claim 9 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and Petkovic et al, US 6185527 (hereinafter Petkovic) and further in view of Friedman et al USPGPUB 20040196989 A1 (hereinafter Friedman).**

Re claim 9, Cooper in view of Zhang and Petkovic fails to teach the system of claim 8, further comprising a component that selects a middle portion of an audio file to mitigate quiet introduction and fades appearing in the audio file. (Friedman [0040]).

Friedman teaches a segment faded-out while another segment is faded in. For example, an audio signal is faded-out (attenuated from full amplitude to silence) quickly (on the order of 0.03 seconds to 0.3 seconds) while the same audio signal is faded-in from an earlier position, such that the end of the faded-in signal is delayed in time.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention selecting a middle portion of an audio file to mitigate quiet introduction and fades in the audio file. Reducing fades and quiet introduction will aid in distinguishing a repeated segment or chorus of a song easier, where quiet portion will have sudden energy changes rather than a gradual change produced by fading.

**12. Claims 15 and 20-22 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Fanty et al US 6535851 B1 (hereafter Fanty).**

Art Unit: 2626

Re claim 15, Cooper in view of Zhang fails to teach the system of claim 1 [[4]], the initial threshold is a normalized Euclidian distance between two fingerprints (Fanty col 5 line 35-45).

NOTE: For prior art purposes, a fingerprint is construed as a vector, where a vector is data from a frame.

Fanty teaches averaged and normalized Cepstral vectors from the left and right compared, where the difference measure is a Euclidean distance. Fanty also teaches the differences measured in step 408 are searched in a left to right manner in order to find local maxima or peaks in the difference measure which are larger than the nearby local minima by more than a threshold amount.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a threshold that is a normalized Euclidian distance between two fingerprints. Using a threshold that is a normalized Euclidian distance between two fingerprints would allow for the detection of data that does not belong in the cluster, such as a verse that doesn't belong in a chorus (i.e. a vector that exceeds or is below a threshold distance).

Re claim 20, Cooper in view of Zhang fails to teach the method of claim 18\_ [[9]], the similarity threshold is a normalized Euclidian distance between two fingerprints (Fanty col 5 line 35-45).

NOTE: For prior art purposes, a fingerprint is construed as a vector, where a vector is data from a frame.



Art Unit: 2626

Fanty teaches averaged and normalized Cepstral vectors from the left and right compared, where the difference measure is a Euclidean distance. Fanty also teaches the differences are searched in a left to right manner in order to find local maxima or peaks in the difference measure which are larger than the nearby local minima by more than a threshold amount.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a threshold that is a normalized Euclidian distance between two fingerprints. Using a threshold that is a normalized Euclidian distance between two fingerprints would allow for the detection of data that does not belong in the cluster, such as a verse that doesn't belong in a chorus (i.e. a vector that exceeds or is below a threshold distance).

Re claim 21, Cooper in view of Zhang fails to teach the method of claim 18, the similarity threshold chosen adaptively based upon the audio file and used to help determine if two fingerprints belong to the same cluster set (Fanty col 5 line 25-45).

Fanty teaches the use of Cepstral coefficients for detecting boundaries in a frame, where the averaged and normalized Cepstral vectors from the left and right are compared and the difference measure is a Euclidean distance. Fanty also teaches the differences are searched in a left to right manner in order to find local maxima or peaks in the difference measure which is larger than the nearby local minima by more than a threshold amount.

NOTE: For prior art purposes, a fingerprint is construed as a vector, where a vector is data from a frame. Also a boundary is construed to be that of the boundary between dissimilar frame containing clusters. Using the measure of a normalized Euclidian distance as a threshold is construed to be functionally equivalent and effective as using an adaptively chosen threshold.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a threshold that is a normalized Euclidian distance between two fingerprints. Using a threshold that is a normalized Euclidian distance between two fingerprints would allow for the detection of data that does not belong in the cluster and distinguishing boundaries from one another, such as a verse that doesn't belong in a chorus (i.e. a vector that exceeds or is below a threshold distance).

Re claim 22, Cooper in view of Zhang fails to teach the method of claim 19[[8]], the clustering operating by considering one fingerprint at a time (Fanty col 13 line 25-36 & Fig. 3).

Fanty teaches chunks of utterance data in sequence where prior utterances can be used. Where a subsequent group of utterances implies one frame of data analyzed at a time.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to consider one fingerprint at a time. Considering one fingerprint at a time would allow for the possibility of analyzing fingerprints in comparison to previous frames in order to establish where a segment such as a chorus repeats.

**13. Claim 16 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Aiken US 6493709 B1 (hereinafter Aiken).**

Re claim 16, Cooper teaches the system of claim 1, wherein a cluster is a group of fingerprints in a set of clusters ([0052] & fig. 7).

However Cooper in view of Zhang fails to teach a set of clusters that lies between the same two gaps or lies between the beginning of the sequence of fingerprints and the first gap in the sequence or lies between the last gap in the sequence and the end of the sequence of fingerprints (Aiken col 3 line 54 – col 4 line 4).

Aiken teaches substrings which are present in the same relative positions in the two strings, which are then stored as a group or displayed as a group. Aiken also teaches multiple substrings or segments from the string of characters having a predetermined length and a beginning position are created, where a predetermined offset or gap between the beginning positions of each consecutive segment is maintained.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention fingerprints in a set of clusters that lies between the same two gaps. Grouping all similar fingerprints within a cluster in between the same gap would allow for the proper segmentation of clusters from one another. Similar fingerprints within a cluster would aid in identifying portions of speech in a song that repeats.

**14. Claims 23 and 24 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Wells et al USPGPUB 20030086341 A1 (hereinafter Wells).**

Re claim 23, Cooper in view of Zhang fail to teach the method of claim 19[[8]], further comprising determining a parameter (D) describing how evenly spread clusters are, temporally, throughout an audio file (Wells [0231]).

Wells teaches thresholds for every value in the fingerprint are based on the observed spread of those values across all songs in the sample set for each value in the fingerprint.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a parameter describing how evenly spread clusters are in an audio file. Having a parameter that describes the spread of a cluster would allow for the calculation of a threshold distance, where clusters can be separated and classified into thumbnails that contain all similar frames based on the distance between frames relative to the threshold distance.

Re claim 24, Cooper in view of Zhang fail to teach the method of claim 23, selecting the set of clusters from which to generate the audio thumbnail based upon at least parameter (D) (Wells [0231]).

Art Unit: 2626

Wells teaches thresholds for every value in the fingerprint are based on the observed spread of those values across all songs in the sample set for each value in the fingerprint.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a parameter that determines which cluster to select to generate an audio thumbnail. Having a parameter that describes the spread of a cluster would allow for the calculation of a threshold distance, where clusters can be separated and classified into thumbnails that contain all similar frames based on the distance between frames relative to the threshold distance.

**15. Claim 34 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Shteyn et al US 6933432 B2 (hereinafter Shteyn).**

Re claim 34, Cooper in view of Zhang fail to teach the method of claim 18, the creating further comprising automatically fading a beginning or an end of an audio thumbnail (Shteyn col 4 line 16-47).

Shteyn teaches a transition piece created by decreasing the volume/loudness at the end of each song (fading-out), then increasing the volume/loudness at the start of the next song (fading-in).

Art Unit: 2626

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention automatically fading a beginning or an end of an audio thumbnail. Fading a song would allow for a smoother transition between segments/frames of a song, such as the transition from song A to song B with a gradual decline in energy levels at the end fade of song A and a gradual incline in energy at the beginning fade of song B.

**16. Claim 35-37 rejected under 35 U.S.C. 103(a) as being unpatentable over Cooper et al USPGPUB 20040073554 A1 (hereinafter Cooper) in view of Zhang USPGPUB 20040064209 A1 (hereinafter Zhang) and further in view of Kanevsky et al US 6434520 B1 (hereinafter Kanevsky).**

Re claim 35, Cooper teaches a log spectrum ([0031]).

However, Cooper in view of Zhang fail to teach the method of claim 18, the generating further comprising processing an audio file in at least two layers, where the output of a first layer is based on a log spectrum computed over a small window and a second layer operates on a vector computed by aggregating vectors produced by the first layer (Kanevsky col 4 line 4-28).

NOTE: For purposes of prior art a layer is construed as a window.

Kanevsky teaches a method based on a discriminative distance measure between two adjacent sliding windows operating on the stream of feature vectors, where two adjacent windows each having a width of approximately 100 frames are

Art Unit: 2626

placed at the beginning of the stream of feature vectors and shifted in time over the audio data stream and the feature vectors of each window are clustered.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention two adjacent sliding windows/layers on a log scale that sum feature vectors. Summing feature vectors on a log scale with several layers/windows would allow for a faster rate of data acquisition as well as a log scaled range of data, which allows the system to process a large range of data into a smaller range of manageable data.

Re claim 36, Cooper in view of Zhang fail to teach the method of claim 35, further comprising providing a wider temporal window in a subsequent layer than a proceeding layer (Herre col 8 line 1-30).

Herre teaches a noise shaping method in the time domain for linear predictive coding, where a filterbank window exhibits only a small overlap between subsequent blocks so that the temporal aliasing effect is minimized. Herre also teaches adaptively selecting a window with a low degree of overlap for critical signals of very transient character while using a wider window type for stationary signals providing a better frequency selectivity.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention a wider temporal window in a subsequent layer than a proceeding layer. Using a wider window in a subsequent layer would allow for the attenuation of

any temporal aliasing effects, which would result in a wider frequency range to select from once the data is inversely transformed.

Re claim 37, Cooper in view of Zhang and Kanevsky fail to teach the method of claim 36, further comprising employing at least one of the layers to compensate for time misalignment (Herre col 8 line 1-30).

Herre teaches a noise shaping method in the time domain for linear predictive coding, where a filterbank window exhibits only a small overlap between subsequent blocks so that the temporal aliasing effect is minimized. Herre also teaches adaptively selecting a window with a low degree of overlap for critical signals of very transient character while using a wider window type for stationary signals providing a better frequency selectivity.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention using a layer to compensate for time misalignment. By overlapping windows, even with a small amount of overlap, a reduced temporal aliasing will be present and time misalignment chances reduced as well as noise.

### ***Conclusion***

17. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. US 20030088412 A1, US 5386493 A, US 20020062209 A1, US 7047194 B1, US 6933432 B2, US 6542869 B1, US 4567606 A, US 6990453 B2, US



Art Unit: 2626

6505160 B1, US 5313556 A, US 7082394 B2, US 5414796 A, US 6963975 B1, US 6606744 B1, US 4241329 A, US 7013301 B2, US 20040260682 A1, US 20030086341 A1, US 20030021472 A1.

18. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael C. Colucci whose telephone number is (571)-270-1847. The examiner can normally be reached on 9:30 am - 6:00 pm, Monday-Friday.

Art Unit: 2626

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Michael Colucci Jr.  
Patent Examiner  
AU 2626  
(571)-270-1847  
[Michael.Colucci@uspto.gov](mailto:Michael.Colucci@uspto.gov)



RICHEMOND DORVIL  
SUPERVISORY PATENT EXAMINER